

MASARYKOVA UNIVERZITA

**Z1069 Statistické metody a zpracování dat**

**I. Úvod, základní pojmy**

**Organizační pokyny**

1. Veškeré potřebné informace jsou uvedeny v ISu (obsah přednášky, doporučená literatura, ...)
2. V ISu budou k dispozici také podklady k přednáškám, vybrané studijní materiály, podklady ke cvičením
3. Cvičení probíhají v programu STATISTICA. Mutilicence MU <https://inet.muni.cz/app/soft/licence>
4. EXCEL, R, ...

- 5. Zakončení předmětu**
- splnění všech požadavků v praktických cvičeních
  - 1. písemný test ( **týden od 9. 11. 2015** ) – řešení praktické úlohy
  - 2. písemný test ( **17. 12. 2015** ) – odpřednášená látka
  - hodnocení: minimálně 50% bodů v součtu obou testů
  - v celkovém hodnocení se promítají i výsledky na cvičeních

**Proč statistika v geografii? (motivace)**

Statistic	Value
Number of Pixels	445200
Mean Value	87.598795
Standard Deviation	22.316273
Minimum Value	0.000000
Maximum Value	255.000000

**Proč statistika v geografii?**

Statistické metody jsou jednou z nedílných součástí **vědecké metodologie**. Cílem vědeckého poznání je **“obecné porozumění”** pozorovaným jevům

**Statistické metody umožňují:**

- zorganizovat data
- popsat je a analyzovat
- formulovat hypotézu
- sestavit model
- otestovat model
- vyslovit zobecnění
- vytvořit, potvrdit / vyvrátit teorii

*Rogerson (2001)*

**STATISTIKA - definice**

**Statistika je vědní obor zabývající se zkoumáním jevů, které mají hromadný charakter.**

Statistika je v určitém smyslu jazykem pro **shromažďování, zpracování, rozbor, hodnocení a interpretaci** hromadných jevů

**Co je typické pro statistiku**

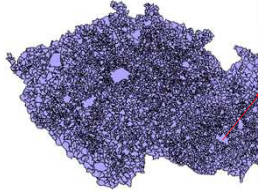
- Zabývá se proměnlivými - **variabilními** - vlastnostmi.
- Pracuje s čísly a vyjadřuje se pomocí čísel - zajímá se především o **kvantitativní stránku** reality (existují však i metody **kvalitativní statistiky**)
- Používá výpočetní techniku k vytváření a správě statistických **databází**, k provádění hromadného **zpracování a analýzy** dat a ke **kommunikaci**.

**Významy pojmu STATISTIKA**

- I. Statistika jako **praktická činnost** - statistická evidence, instituce, ročenky meteorologických pozorování atd.
- II. Statistika jako **vědní disciplína** - popisná a matematická (induktivní) statistika, aplikované vědy (ekonometrie, chemometrie atd.), vědy se silným statistickým základem: klimatologie, hydrologie, sociologie, psychologie, demografie aj.

## Vymezení základních pojmů I

**Hromadné jevy:** přírodní či společenské jevy, které jsou výsledkem působení velkého množství příčin, jejich vlastnosti se neprojevují v jednotlivých jevech, ale jen v souboru a to prostřednictvím řady náhod.



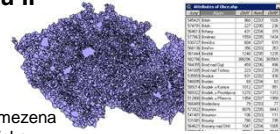
Adresy of (obozna)	2007	2007	2007
142421 Brno	966	1202	1373
142422 Brno	227	1206	278
142423 Brno	427	1204	195
177923 Brno	1854	1206	1434
150727 Brno	568	1202	1572
148146 Brno	364	1203	393
101444 Brno	1346	1202	1579
155795 Brno	38026	1206	30369
154100 Brno	455	1206	480
141805 Brno nad Tuhou	227	1203	228
150903 Brno	371	1202	450
148095 Brno	63	1204	61
100103 Brno-Líšeň	1872	1202	1901
100122 Brno-Líšeň	1374	1202	1372
172828 Brno-venkov	394	1202	1960
156883 Brno	75	1203	77
172822 Brno	8074	1206	8441
141440 Brno	16	1203	111
151001 Brno	786	1202	796
194621 Brno nad Ohří	1347	1204	1365

Řada jevů, které v geografii studujeme pomocí statistických metod, má povahu jevů náhodných – tzv. **stochastických** (hydrologické jevy či meteorologické jevy).

Stochastické vs. **deterministické** jevy, procesy

## Vymezení základních pojmů II

**Statistická jednotka:** je to určitý jev či prvek, který je předmětem statistického šetření a pro který se zjišťují údaje



Statistická jednotka musí být přesně vymezena na počátku vlastního šetření a to z hlediska **věcného, časového, prostorového**.

**Statistický znak:** je to určitá vlastnost statistické jednotky, kterou se snažíme postihnout.

**Statistický soubor:** skupina statistických jednotek stejného druhu (věcně, prostorově a časově vymezených), které jsou předmětem statistického zkoumání. Každý z prvků je statistickou jednotkou.

Prvky tvořící statistický soubor mají určité společné vlastnosti - tzv. **identifikační znaky** - umožňující určit, zda prvek do daného statistického souboru patří nebo nepatří (**vymezují** statistický soubor).

Z hlediska cílů statistického zkoumání sledujeme na prvcích statistického souboru jednu nebo více vlastností - **sledované znaky**.

## Vymezení základních pojmů III

Statistické znaky lze dělit na znaky **prostorové, časové a věcné**.

Věcné znaky se dělí na znaky **kvantitativní** a **kvalitativní**

Kvalitativní znaky mohou být **alternativní** a **množné**

Kvantitativní znaky dělíme nejčastěji na znaky **spojité** a **diskrétní**.

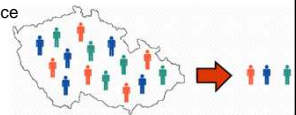
Statistické znaky můžeme získat přímo – (např. **měřeními**) a nebo **nepřímě** (výpočtem). Tyto potom nazýváme znaky odvozenými.

Podle škály, na které znaky zjišťujeme je dělíme na znaky **nominální, ordinální, poměrové, intervalové** (viz dále)

## Vymezení základních pojmů IV

**Základní statistický soubor** - populace

**Výběrový statistický soubor** je podmnožinou základního souboru. Je vytvořen ze statistických jednotek, vybraných podle určitého hlediska.



**Reprezentativní výběr:** pokud zkoumaný výběr dobře odráží strukturu celého zkoumaného souboru, nazýváme jej reprezentativním výběrem.

Statistický soubor **jednorozměrný, vícerozměrný**

**Rozsah statistického souboru:**

$N$  – rozsah základního souboru

$n$  – rozsah výběrového souboru

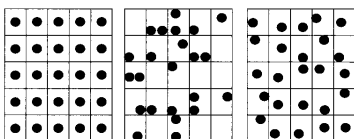
## Vymezení základních pojmů V

**Výběrová šetření**

S výběrovými soubory pracujeme, protože např. z důvodů omezeného času či prostředků nemůžeme vyšetřit všechny jednotky populace či počet statistických jednotek v populaci není konečný.

**Náhodný výběr:** každý prvek populace má stejnou možnost být zařazen do výběru

Výběr mechanický (systematický), oblastní (stratifikovaný), více-stupňový, ...



Příklady systematického, náhodného a stratifikovaného náhodného výběru

## Popisná statistika

**Popisná (deskriptivní) statistika** se zabývá uspořádáním souborů, jejich popisem a účelnou sumarizací.

Jak mohou být tyto jevy jednoduše popsány (charakterizovány, sumarizovány).

Existují dvě základní možnosti, které se vzájemně doplňují:

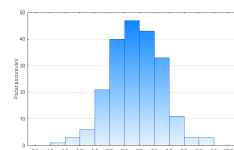
• **numerické metody** – jedním nebo několika málo čísly lze vystihnout určité vlastnosti jevu. Jsou přesnější a objektivnější.

$$\bar{x} = 0,5$$

$$x_{\min} = -3,6$$

$$x_{\max} = 3,0$$

• **grafické metody** – sestrojení vhodného typu grafu. Jsou názornější a umožňují vystihnout vztahy.



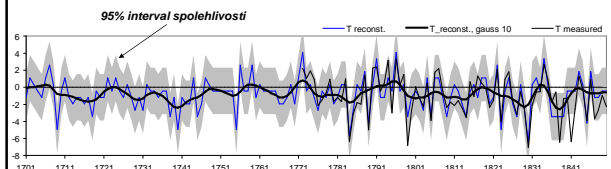
## Induktivní statistika

**Induktivní (matematická) statistika** se vyvinula z popisné statistiky a jejím základem je **teorie pravděpodobnosti**.

Matematická statistika zkoumá soubory nepřímo prostřednictvím výběrů.

Induktivní statistika se zabývá metodami jak poznatky **přenášet** a umožňuje z pozorovaných dat vytvářet **obecné závěry** s udáním stupně jejich spolehlivosti.

Výpočet stupně spolehlivosti závěrů je však objektivní, neboť je založen na poznacích teorie pravděpodobnosti a nezávisí na subjektivním názoru hodnotitele.



## Geografická data a jejich specifika

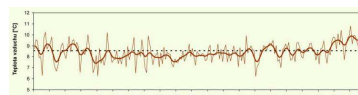
- **Zdroje geografických dat** – primární, sekundární
- **Prostorový aspekt** – statistika prostorově lokalizovaných dat (**geostatistika**)



Dr. John Snow



- **Časový aspekt**



Místa výskytu onemocnění cholery v Londýně v r. 1854

## Geografická data a jejich specifika

Dělení geografických dat podle použité škály měření:

- **Nominální** (kategorie využití země)
- **Ordinální** (řád vodního toku, stupnice síly větru)
- **Intervalová** (teplota vzduchu) nula = data
- **Poměrová** (množství srážek, délka vodního toku) nula = neexistence jevu

## Typy geografických dat

**Nominální data** – hodnota představuje konkrétní kategorii či třídu a vyjadřuje její označení (jméno), kategorie se nesmějí překrývat – jsou disjunktní. Každý objekt je zařaditelný alespoň do jedné kategorie, žádný nespadá do více než jedné. Čísla, která označují kategorie jsou pouze symboly a nelze s nimi provádět aritmetické operace. V nejjednodušší podobě mají binární charakter a lze je pouze porovnávat.

**Ordinální data** – data, která lze seřadit do uspořádané posloupnosti podle určitého kritéria. Je známé pořadí kategorií, rozdíl však nemá smysl. Např. řád vodního toku, třída silnice, bonita půdy atd.

## Typy geografických dat

**Intervalová data** – umožňují provádět i odečítání mezi kategoriemi definovat rozdíl mezi kategoriemi. Teplota vzduchu. Stupnice většinou nezačíná nulou. Poměr dat závisí na zvolených jednotkách.

**Poměrová data** – vedle rovnosti, uspořádání a odečítání umožňují také dělení. Nula vyjadřuje neexistenci jevu – objem, délka ...

## Základní vyjadřovací prostředky ve statistice

- Statistické tabulky
- Statistické grafy

Tabulky – složené z buněk, přehledné, nezávislé na textu

Tab. 1 Základní statistické charakteristiky teploty vzduchu [°C] na vybraných stanicích za období 1961-2000

Charakteristika	Stanice	
	Praha- Klementinum	Strání
průměr [°C]	9,205	7,673
rozptyl <sup>1)</sup>	0,679	0,403
rozsah souborů	120	30
směrodatná odchylka [°C]	0,821	0,624
F	1,687	-
F (kritické)	1,699	-

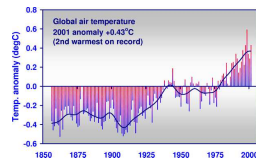
Pramen: ČHMÚ  
Vysvětlivky: 1) ...

- nadpis
- záhlaví
- legenda
- pramen
- poznámky
- vysvětlivky

## Statistické tabulky

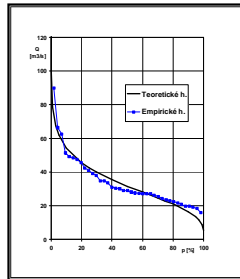
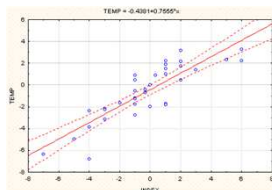
- Záhlaví a legenda mají obsahovat měrné jednotky
- Tabulka má vyplněna všechna políčka
- Smluvené znaky pro políčka bez číselného údaje
  - – údaj se nevyskytuje
  - x – údaj není možný z logických důvodů
  - 0 – hodnota je menší než polovina nejmenší měrné jednotky
  - – údaj nelze spolehlivě zjistit

## Metody grafického znázornění geografických jevů



Motto:

Jeden obrázek je za tisíce slov

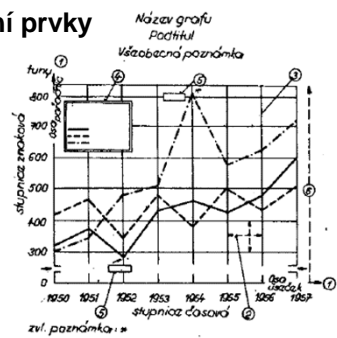


Cílem grafického znázornění je podat rychlou a srozumitelnou informaci o studovaném jevu či o vzájemném vztahu více jevů.

## Graf a jeho základní prvky

Graf – kresba provedená podle předem dohodnutých pravidel, která znázorňuje kvalitativní či kvantitativní znaky.

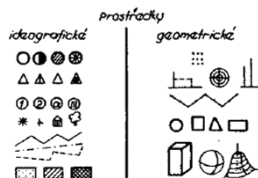
1. stupnice
2. grafický interval
3. síť
4. klíč
5. vysvětlivka
6. délka stupnice



**Grafický obraz** – soubor grafických prostředků, pomocí kterých na základě dohodnutého výkladu jejich smluveného významu sestrojíme graf

**Grafický výklad** – soubor zásad, podle kterých interpretujeme (čteme) příslušný graf).

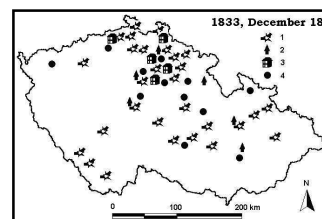
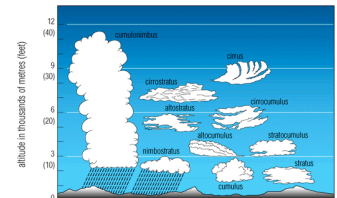
## Dělení grafických prostředků podle významu



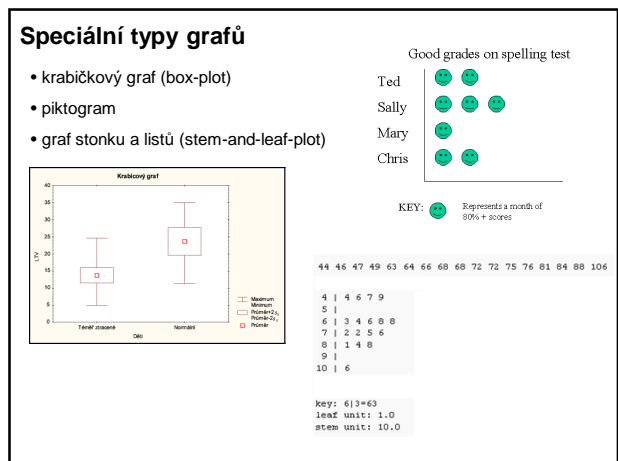
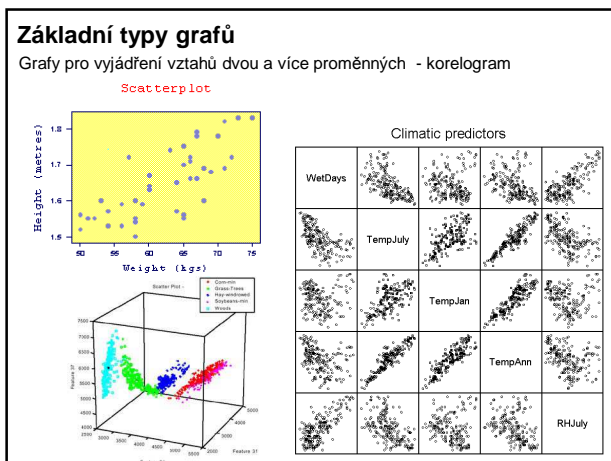
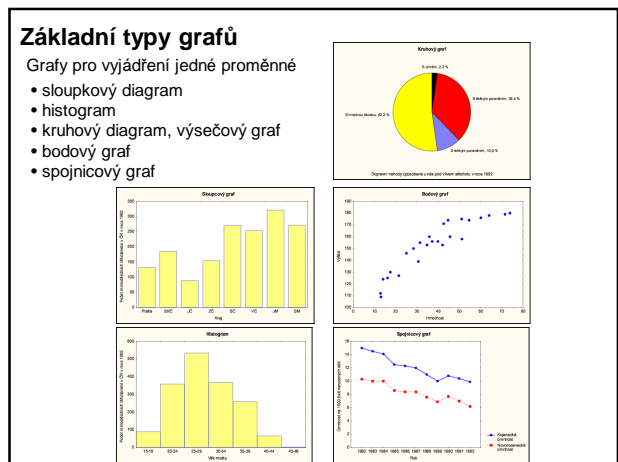
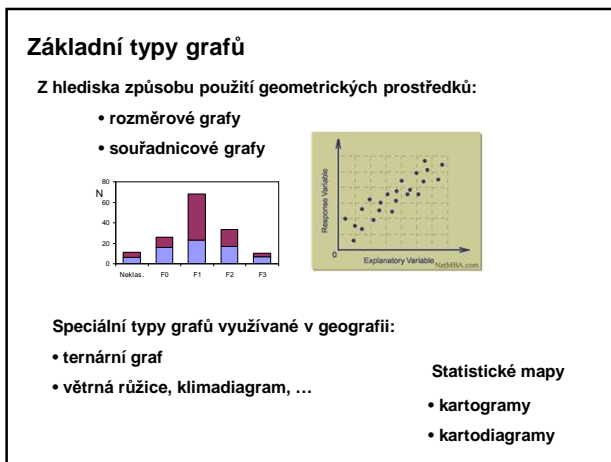
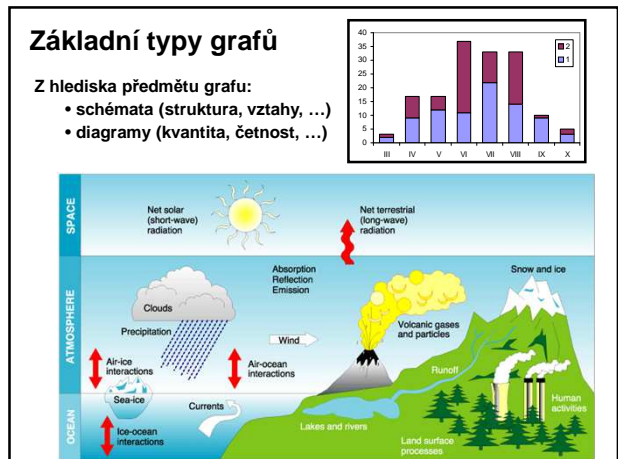
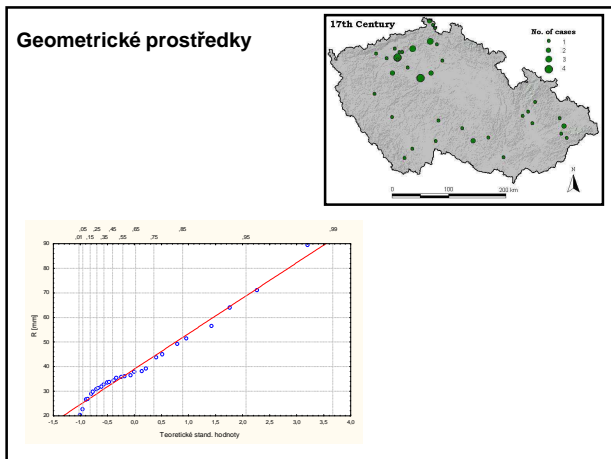
**Ideografické** – mají kvalitativní význam a v grafu fungují jako znaky (klasifikační, identifikační). Jejich tvar a rozměry slouží pouze k jejich odlišení, nemají kvantitativní význam (písmena, číslice, symboly, geometrické obrazce, šrafura, barva, druhy čar apod.).

**Geometrické** – mají vždy kvantitativní význam, často však také slouží ke kvalitativnímu odlišení statistických jednotek (body, úsečky, obrazce).

## Ideografické prostředky



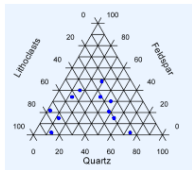
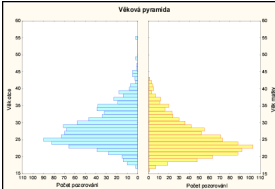
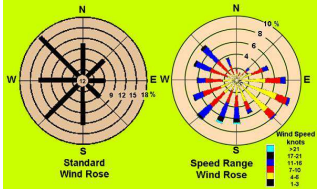
- 1 – windbreakage
- 2 – damage on buildings of lesser extent
- 3 – destroyed buildings
- 4 – damage without specification





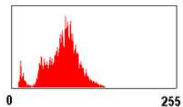
### Speciální typy grafů využívané v geografii:

- ternární graf
- „strom života“
- větrná růžice
- klimadiagram

### Analýza grafů

Všimáme si základního tvaru a odchylek od něho

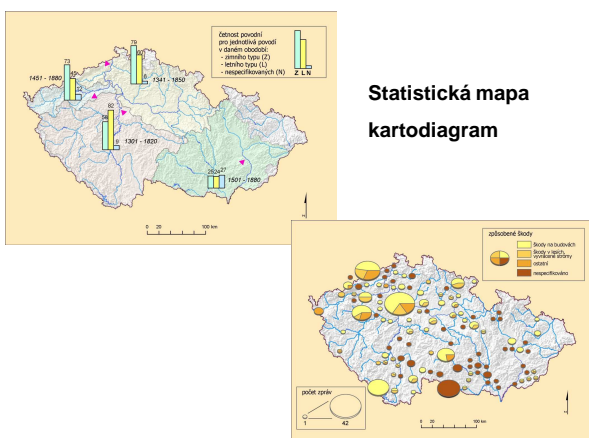


**U tvaru grafu hodnotíme:**

- zhuštění – místa největší četnosti hodnot
- shluky – existence jednoho či více shluků hodnot
- mezery – existence intervalů či oblastí bez hodnot
- odlehle hodnoty – existence údajů podstatně rozdílných od ostatních hodnot
- extrémní hodnoty – poloha min a max hodnot v grafu
- tvar rozdělení – jak ho lze popsat – symetrie, počet vrcholů

Volba vhodného typu grafu musí zohledňovat typ zobrazované proměnné (spojitá či diskrétní)

### Statistická mapa kartodiagram



**Legenda:**

- bodové poměry pro jednotlivá poměry v dané oblasti
- směrný typ (Z)
- směrný typ (L)
- nespecifikačních (N)
- Z, L, N

**zobrazované body**

- body roztroušené
- bodové poměry
- osy
- speciální

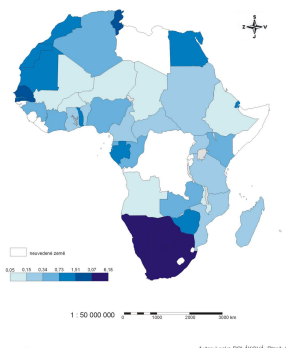
**počet znaků**

1 42

### Příklady kartogramu

Grafické vyjádření intenzity jevu

**POČET OSOBNÍCH POČÍTAČŮ NA 100 OBYVATEL**  
v kontinentální Africe a na Madagaskaru v roce 2000



**Z-skóre**

- 1,832
- 1,892 - -1,693
- 1,693 - -0,267
- 0,267 - 0,176
- 0,176 - 0,484

1 : 50 000 000

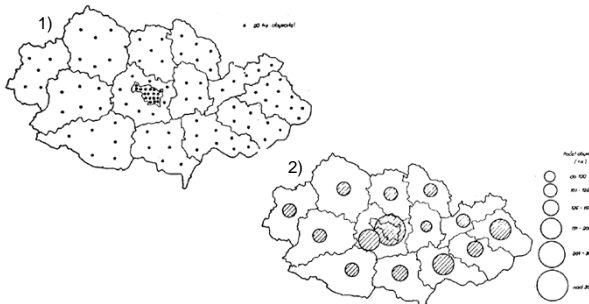
GL BUS

Autar, Lenka POLÁČKOVÁ, Praha, 2003  
Vydáno nakladatelstvím GL BUS v Praze 2005

### Grafické znázornění prostorové diferenciace intenzity jevů:

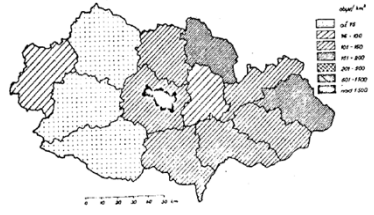
• Absolutní metoda:

- 1) bodová – počet grafických prvků je úměrný velikosti jevu
- 2) značková – velikost grafického prvku je úměrná velikosti jevu

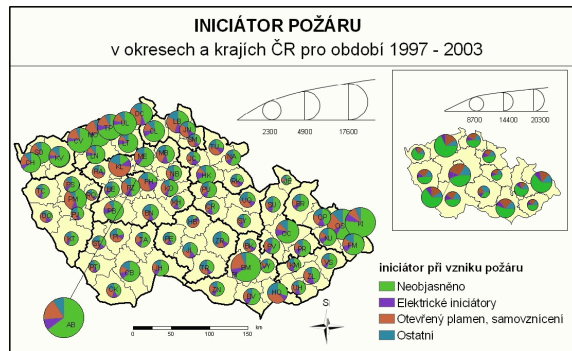


### Grafické znázornění prostorové diferenciace intenzity jevů:

Relativní metoda: barva, šrafování

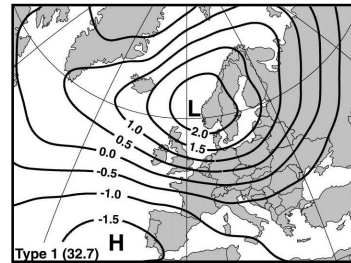
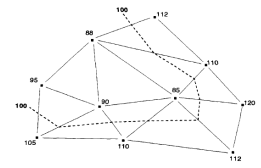


Grafické znázornění prostorové diferenciace intenzity a struktury jevů



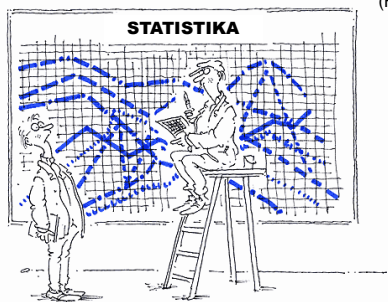
Konstrukce izoliní - interpolace:

Konstrukce spojitých polí  
(viz. Geostatistika)



Statistika se těší pochybnému vyznamenání tím, že je nejvíce nepochopeným vědním oborem

(H. Levinson)



VÝBORNĚ JISTĚNĚTE PRĚNĚ, KDYŽ TĚMU NIKDO NEBUDE ROZUMĚT.